Theor Appl Genet (2002) 105:1151–1158
DOI 10.1007/s00122-002-0967-4

L.L. Georgi · Y. Wang · D. Yvergniaux · T. Ormsbee
M. Iñigo · G. Reighard · A.G. Abbott

Construction of a BAC library and its application to the identification of simple sequence repeats in peach [*Prunus persica* (L.) Batsch]

Received: 11 October 2001 / Accepted: 8 March 2002 / Published online: 4 September 2002

Abstract The Rosaceae contains many economically important crop species, but their genomes are not well characterized, and comparative genetic mapping lags well behind that of other families. To facilitate genome comparisons and gene discovery in the Rosaceae, we have begun the development of genomic resources for peach as the model genome for this family. First, we developed a simplified, cost-effective method for constructing BAC libraries, particularly appropriate for plant species of relatively minor economic importance. Second, we used the library to investigate the abundance and local distribution of simple sequence repeats (SSRs) in peach. Our results indicate that microsatellite loci are locally much more highly abundant than previously estimated, and BAC sequencing results suggest that microsatellite repeats are not randomly distributed within gene-containing regions of the peach genome. This makes it relatively easy to identify SSRs in peach by hybridization to BAC clones, and even by random sequencing of BAC clones, not known a priori to contain SSRs.

Keywords Microsatellite · Retrotransposon · Rosaceae · Bacterial artificial chromosome

Introduction

The Rosaceae is economically important in temperate regions. Important fruit-producing crops in the family include members of the following genera: *Malus* (apple), *Pyrus* (pear), *Rubus* (raspberries/blackberries), *Fragaria* (strawberries) and *Prunus* (stone fruits). Of these genera, *Prunus* is the most diverse, with a large array of economically important fruit (peach/nectarine, apricot, plum, sweet and sour cherry), nut (almond) and lumber (black cherry) tree species. Additionally, the Rosaceae contains a wide variety of ornamental plants including roses, flowering cherry, crabapple, quince and pear.

Despite the collective importance of Rosaceae crops, structural and functional genomic characterization of the member species lags well behind that of other families. Several factors have contributed to this lack of progress. First, although these crops collectively are very important worldwide, individual crops are of more regional importance. Additionally, the majority of Rosaceae fruit crops are trees, and linkage mapping is difficult due to the field space required and the length of time to flowering and fruiting (at least 2 years). Yet another difficulty is the polyploid nature of many of the crops: for example, cultivated strawberry is octoploid, sour cherry is tetraploid, and some plum species are hexaploid (Okie and Weinberger 1996)

In other plant families (most notably the Poaceae), comparative mapping has accelerated the genetic understanding of the individual member crops via the utilization of physical mapping resources in rice (Delseny et al. 2001). Within the Rosaceae, however, comparative mapping has only begun among *Prunus* species (peach, almond, sour cherry), and *Prunus* maps have not yet been compared to maps of other Rosaceous species. Comparative mapping and gene discovery in the Rosaceae are currently limited by the low marker density of genetic linkage maps of member crops and the low level of marker transfer between maps of different species. Recent studies have demonstrated the utility of certain types of molecular markers, in particular simple sequence repeat (SSR) markers or microsatellites, which are abundant in *Prunus* genomes and exhibit sufficient polymorphism to allow a comparison of maps among populations and species (Sosinski et al. 2000). However, most Rosaceae maps do not contain many mapped SSRs, and only recently have

Communicated by D.B. Neale

L.L. Georgi (✉) · D. Yvergniaux · T. Ormsbee · M. Iñigo
A.G. Abbott
Department of Genetics and Biochemistry, Clemson University,
Clemson SC 29634, USA
e-mail: georgil@clemson.edu
Tel.: +864-656-3060, Fax: +864-656-6879

Y. Wang · G. Reighard
Department of Horticulture, Clemson University,
Clemson SC 29634, USA

Table 1 Current genome maps of Rosaceae species

Species	Reference	Map specifics	
		No. of markers	Map size
Peach	Chaparro et al. 1994	83	396 cM
	Rajapakse et al. 1995	64	465 cM
	Lu et al. 1998	157	1,400 cM
	Dirlewanger and Bodo 1994		
	Dirlewanger et al. 1998	249	712 cM
	Shimada et al. 2000	83	960 cM
Peach/Almond	Foolad et al. 1995	100	800 cM
Cherry	Wang et al. 1998	161	508 cM
Apricot	Hurtado Ruiz 1999	115	570 cM
		64	448 cM
Prunus (general)	Joobeur et al. 1998	213	491 cM
Almond	Viruel et al. 1995	69	394 cM
		93	393 cM
	Joobeur et al. 2000	126	415 cM
		99	416 cM
Apple	Hemmat et al. 1994	300	1,120 cM
	Maliepaard et al. 1998		984 cM
Strawberry	Davis and Yu 1997	88	450 cM
Rose	Rajapakse et al. 2001	171	902 cM

laboratories begun to invest heavily in developing this marker base (Cipriani et al. 1999; Joobeur et al. 2000; Sosinski et al. 2000). In contrast, a number of Rosaceae maps have been developed with other marker types, including RFLPs, RAPDs and AFLPs (Table 1). These maps are marker dense, they cover substantial portions of the genome and, in the case of AFLP markers, they can be generated quickly and efficiently.

The process of comparative genomics and gene discovery in the Rosaceae could be greatly accelerated by the development of a well-characterized genomic resource for a Rosaceae model species. This resource would consist of three fundamental units: a physical map, integrated genetic marker maps, and mapped ESTs (Expressed Sequence Tags). A physical map would serve as the foundation on which the genetic marker and EST information could be layered. Placement on a common physical map of molecular markers from genetic maps of the various species would enable alignment of these maps and facilitate the identification of genomic regions containing important loci, thus speeding the process of gene tagging and discovery. Peach is a logical candidate model species for the Rosaceae, because it is diploid with a relatively small genome [haploid size of 300 Mb (Baird et al. 1994)] and, though a tree species, it has a comparatively short juvenile period (2 to 3 years). For these reasons, we have produced a BAC library of the peach genome and are currently constructing contigs containing mapped molecular markers from various *Prunus* species as a way of initiating construction of a physical map, while simultaneously integrating information from genetic maps of various rosaceous species. Here we report a significantly simplified procedure for BAC library construction, and the utilization of the library for assigning a range of mapped molecular markers to physical map contigs and for examining the abundance, types, and distribution of SSRs in the peach genome.

Materials and methods

BAC library construction

High-molecular-weight (50–300 kb) DNA was extracted in solution from leaves of the peach rootstock Nemared by a method similar to that outlined in ten Lohuis et al. (1993). Specifically, 25 g fresh of leaf tissue was homogenized in 200 ml of ice cold 1 × HB (0.5 M sucrose, 10 mM Tris base, 80 mM KCl, 10 mM EDTA, 1 mM spermidine, 1 mM spermine, pH 9.4) plus 0.15% beta-mercaptoethanol, for 30 s in a Waring blender. This material was squeezed through two layers of cheesecloth and one layer of Miracloth (Calbiochem) into a cold 250-ml centrifuge bottle (it is easier to strain through cheesecloth first, then through Miracloth). After the addition of 5 ml (4:1) of 1 × HB:Triton X-100, the bottle containing the mixture was incubated on ice for 20 min. Nuclei were recovered by centrifugation at 3,300 rpm in a Sorvall GSA rotor for 20 min at 4 °C and the supernatant was discarded. The nuclei were resuspended in 1 ml of ice cold 1 × HB plus 0.5% Triton X-100 using a paint brush and transferred into a 50-ml Oakridge tube with an additional approximately 30 ml of the HB-Triton solution. Nuclei were recovered by centrifugation at 3,300 rpm in a Sorvall HB4 rotor for 15 min at 4 °C (this washing step may be repeated). The washed nuclei were suspended as before in 1 ml of 1 × HB plus Triton, then lysed by the addition of an equal volume of lysis buffer (50 mM Tris pH 8, 10 mM EDTA, 2% Sarkosyl). Cesium chloride was dissolved in the lysate (0.97 g of CsCl per ml of lysate). The solution was transferred to a Corex tube, centrifuged in a Sorvall SS34 rotor at 8,000 rpm for 20 min at 4 °C to remove protein, and subsequently transferred to an ultracentifuge tube and centrifuged for 2.5 days at 20 °C and 175,000 g. No ethidium bromide was used in the gradient. Fractions were collected from the gradient using cut-off pipette tips to avoid shearing the DNA. DNA-containing fractions were identified by diluting 2 µl of each fraction in 3 µl of water and electrophoresis on a 0.8% agarose gel. Positive fractions were dialyzed against TE (10 mM Tris, 1 mM EDTA pH 8) to remove Cesium chloride. DNA was concentrated by ethanol precipitation, hooked out with an end-sealed, bent Pasteur pipette, and resuspended in TE. This solution was stored at 4 °C. We have found high-molecular-weight DNA prepared and stored in this manner to be stable for at least 2 years.

DNA was partially digested with *Hin*dIII by incubating 500 ng of DNA (estimated spectrophotometrically) with 0.1 units of enzyme for 20 min at 37 °C, then incubating for 15 min at

65 °C to inactivate the enzyme. The digested DNA was electrophoresed on a conventional mini-agarose gel (0.5% SeaKem GTG agarose in 1 × TAE, 40 mM Tris-acetate,1 mM EDTA, pH 8.0) at 100 V for 1 h at room temperature, using autoclaved running buffer. Before use, the gel apparatus was cleaned with 2% abSolve (NEN Research Products) and rinsed with sterile distilled water. Marker lanes were cut from the gel and stained with ethidium bromide; notches were cut in the stained gel to mark the migration of uncut lambda DNA (approximately 50 kb). The stained and unstained parts of the gel were aligned and a gel slice containing fragments larger than 50 kb was cut from the unstained part of the gel. These fragments were electroeluted from the gel into dialysis tubing for 2 h (Sambrook et al. 1989), concentrated by Centricon 100 (Millipore) spin dialysis, and ligated into pBeloBAC11. Ligation reactions contained approximately 40 ng of vector in an estimated five-fold molar excess over insert DNA, and 4.2 units of T4 DNA ligase (Promega) in a 100-µl reaction volume, and were incubated overnight at 16 °C. The ligations were then either spin dialyzed (Centricon 100) or "dialyzed" by pipetting into depressions in autoclaved 0.1 M glucose solidified with 1% SeaKem LE agarose and incubating on ice for 90 min (Jose Luis Goicoechea, personal communication). Ligation products were introduced into *Escherichia coli* strain DH10B (Electromax, Gibco BRL) by electroporation using a BRL Cell-Porator and Voltage Booster set at 320 V, 330 µF, Impedance Low Ohms, Fast charge rate and 4 K Ohms resistance. The cells were added to 400 µl of SOC and incubated, shaking, at 37 °C for 1 h; then a portion of the transformation was plated on selective media (LB plus chloramphenicol, IPTG and X-Gal). Selected clones were grown up and their DNA extracted (see method below) and electrophoresed undigested or digested with either *Hin*dIII or *Not*I, to verify the presence of large inserts prior to transforming and plating large numbers of clones for the library. Plates were grown overnight at 37 °C, then held for an additional day in the dark at room temperature to permit the color to develop sufficiently for detection by the robot. The resulting clones were robotically arrayed by a Genetix Q-Bot into 384-well plates and stored at –80 °C (Choi et al. 1995).

Hybridization

Clones from the peach BAC library plates were robotically arrayed in duplicate on two and one half 22 × 22 cm Hybond N+ (Amersham) filters for hybridization. One library filter set was sequentially hybridized with tomato cDNA clone pTC11 (Mao et al. 2001), peach cDNA clone pP3Mi22, and peach AFLP-derived STS clone EAA/MCAT10 (Lu et al. 1999). Additional filter sets were also hybridized with peach molecular markers (cDNA, RFLP and AFLP, see Table 2). Probes were labelled with α^{32}P dCTP by the method of random priming (Feinberg and Vogelstein 1983). Additionally, 26 randomly selected peach BAC clones were manually replicated onto 100 cm-diameter round Hybond N+ filters for colony hybridization. These filters were probed separately with $(GT)_{10}$, $(CT)_{10}$ and $(AGG)_{10}$ oligonucleotide probes end-labelled with γ^{33}P-ATP using T4 polynucleotide kinase (Promega). After 2 h of prehybridization in 7% SDS, 0.25 M sodium phosphate buffer (pH 7.4), hybridizations were carried out overnight in the same solution. Prehybridization and hybridization temperatures were 60 °C (tomato cDNA probe), 65 °C (peach probes) or 55 °C (oligonucleotide probes). Filters were then washed twice at the hybridization temperature, each time for 20 min, in 2 × SSC, 0.1% SDS; filters hybridized with peach probes were washed two additional times (15 min each) in 0.3 × SSC, 0.1% SDS. Hybridization signals were detected autoradiographically using Kodak XAR film.

BAC DNA prepared from positive clones was digested with *Hin*dIII, electrophoresed on 0.8% SeaKem LE agarose and stained with ethidium bromide. Southern transfer of the DNA to Hybond N+ membranes was performed using the manufacturer's Alkaline Transfer Protocol. Southern blots were hybridized with the appropriate probe as described for colony blots.

BAC DNA extraction

Peach BAC DNA was extracted by a modified alkaline-lysis procedure (Jose Luis Goicoechea, personal communication): clones were grown overnight in 5 ml of LB broth and harvested by centrifugation in a Beckman GPR centrifuge, GH-3.7 swingout rotor for 10 min at 4 °C. Pelleted cells were resuspended in 0.3 ml of GTE (50 ml of glucose, 25 mM Tris pH 8, 10 mM of EDTA pH 8) containing 0.1 mg/ml of RNase A, transferred into microfuge tubes, and lysed by adding 0.3 ml of 0.2 N NaOH/1% SDS and incubating at room temperature for 2.5 min. The lysates were neutralized by adding 0.3 ml of 3.0 M potassium acetate, pH 4.8. Following a 5-min incubation on ice, cellular debris was removed by centrifugation in a microfuge for 5 min at room temperature. BAC DNA was precipitated by adding 0.7 vol of isopropanol. Larger-scale BAC DNA extractions (50–500 ml culture volume) were additionally purified on cesium chloride density gradients using a Beckman TL100 ultracentrifuge.

Subcloning and sequencing

For sequencing, BAC DNA was digested with *Hin*dIII, and the resulting fragments were ligated into pUC119 and transformed into *Escherichia coli* strain DH5α by calcium/heat shock (Sambrook et al. 1989). Plasmid DNA for sequencing was prepared using a protocol very similar to the one described above for BAC DNA, followed by precipitation from 0.4 M NaCl, 6.5% PEG. Subclones were sequenced using ABI's Dyedeoxy terminator cycle sequencing kit and an ABI 373A or 377 DNA sequencer, or Amersham's Thermosequenase kit and a Licor 4200L sequencer. Sequences were assembled using Sequencher 3.1.1 software (GeneCodes Corp.).

SSR amplification

For selected SSR loci identified in BAC sequences (see Table 3), primers were designed to complement flanking sequences using the program Primer3 (Rosen and Skaletsky 1998). SSR loci were amplified from 4 ng of peach genomic DNA in 10-µl reactions containing 10 mM of Tris–HCl (pH 8.3), 50 mM of KCl, 1.5 mM of $MgCl_2$, 200 µM of each dNTP, 0.2 µM of each primer, and 0.5 U of *Taq* polymerase. After an initial denaturation at 94 °C for 4 min, the reactions were subjected to 32 cycles of 94 °C for 30 s, 50 °C for 30 s and 72 °C for 30 s, followed by a final extension at 72 °C for 5 min. Each locus was evaluated for amplification and polymorphism in peach rootstocks Lovell and Nemared, by electrophoresing the PCR products on 3% (3:1 Nusieve GTG:Seakem LE) agarose gels and staining with ethidium bromide. If no differences were resolvable on agarose gels, the amplifications were repeated using a radiolabelled primer, as described in Sosinski et al. (2000)

Results

BAC library construction

To initiate genomic resource development in peach, we constructed a peach BAC library using pBeloBAC 11 and large genomic DNA fragments isolated by a simplified CsCl gradient procedure, with conventional gel electrophoresis for size selection of partially digested DNA. Once this method was optimized for peach, actual library construction was completed in 2 weeks. The Nemared peach library consists of 44,160 BAC clones. Insert size averaged 50–70 kb and ranged from 50 kb to over 100 kb (Fig. 1). Assuming a haploid genome size of

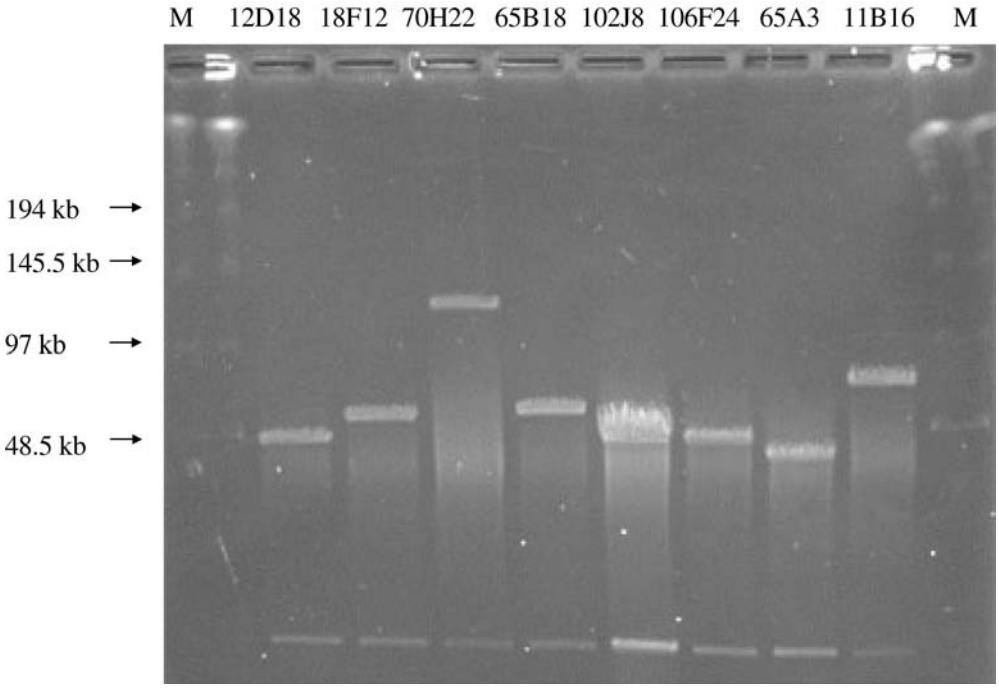

Fig. 1 Ethidium bromide-stained pulsed-field gel of eight peach BAC clones digested with *Not*I. Clones 11B16,12D18, 65B18 and 70H22 hybridized with peach AFLP probe AT/CTA; clone 18F12 hybridized with AFLP probe AT/CAC; and clones 65A3, 102J8 and 106F24 hybridized with AFLP probe TT/CCA

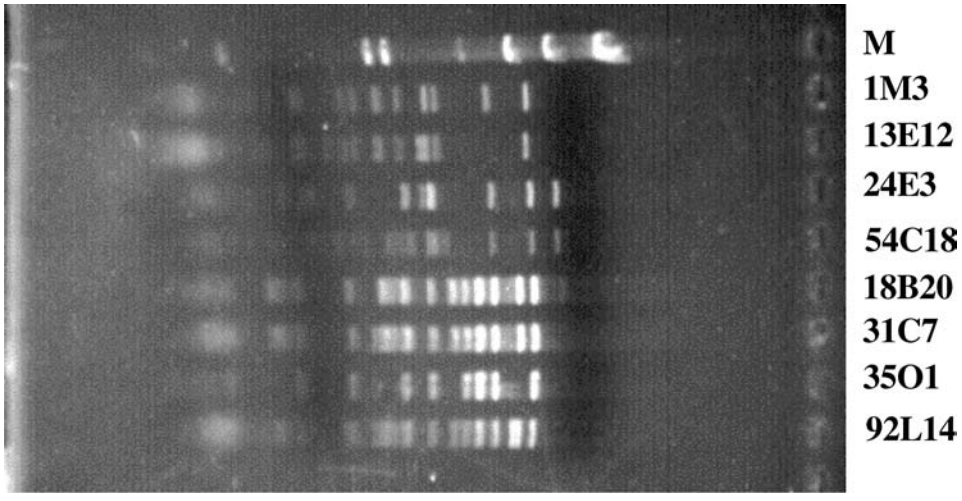

Fig. 2 Peach BAC clones selected by hybridization with peach cDNA clone P3Mi22 (BAC clones 1M3, 13E12, 24E3 and 54C18) or tomato cDNA clone pTC11 (BAC clones 18B20, 31C7, 35O1 and 92L14), digested with *Hin*dIII; the molecular-weight marker (*lane 1*, *M*) is phage lambda DNA digested with *Hin*dIII. Ethidium bromide-stained agarose gel

300 Mb (Baird et al. 1994), this theoretically represents a greater than eight-fold coverage of the peach genome. Hybridization of the library with RFLP, AFLP and cDNA probes yielded from 1 to 14 clones per probe, with an average of 4.6 library clones detected by a given probe sequence (Table 2). Some of these probes, AC37 for example, may hybridize to more than one genomic region (Joobeur et al. 1998), but others appear to be single-copy. When co-hybridizing sets of BAC clones were digested and electrophoresed, multiple co-migrating bands were shared within a set but not between sets of clones (Fig. 2). This provided additional evidence of overlapping sequences between clones hybridizing with the same probe and permitted their ordering in contigs as well as the estimation of contig sizes. The average size of nine such non-singleton contigs (each comprising 2 to 10 co-hybridizing BACs) was estimated to be 69.7 kb.

Occurrence of SSRs in BAC clones

SSR abundance in localized genomic regions of peach was evaluated by Southern hybridization of end-labelled oligonucleotide probes to peach BAC clones. Colony hybridization of 26 randomly chosen BAC clones yielded 5, 10 and 20 clones positive for $(AGG)_n$, $(GT)_n$ and $(CT)_n$, respectively. Only four clones were negative for all three probes, indicating that over 80% of the peach BACs contained at least one SSR. Two of the non-hybridizing clones appeared to contain several copies of a 9.6-kb tandem repeat. Hybridization of the appropriate repeat probe to Southern blots of *Hin*dIII-digested BACs positive for either the $(CT)_n$ or $(GT)_n$ probes revealed that each BAC tested had one to four fragments containing (CT) repeats, and one to four fragments containing (GT) repeats (Fig. 3).

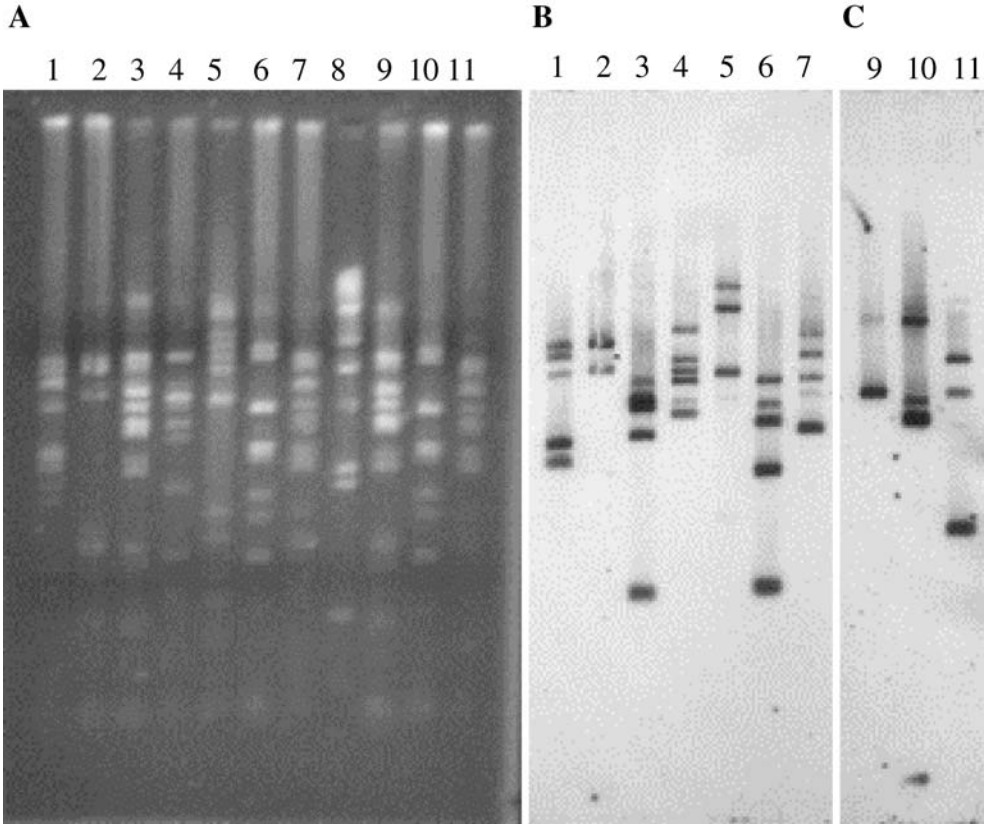

Fig. 3A–C Randomly chosen peach BAC clones hybridizing with (CT) and (GT) repeats. **A** Ethidium bromide-stained gel of BAC DNAs digested with *Hin*dIII; the molecular-weight marker (*lane 8*) is phage lambda DNA digested with *Hin*dIII. **B** Southern blot of *lanes 1–7* of the gel, probed with a $(CT)_{10}$ oligonucleotide end-labelled with γ^{33}P-ATP. **C** Southern blot of *lanes 9–11* of the gel, probed with a $(GT)_{10}$ oligonucleotide, end-labelled with γ^{33}P-ATP. Note that *lanes 9–11* repeat *lanes 3, 6, and 7*

Table 2 Molecular genetic markers assigned to physical map contigs

Probe type	Probe	Genetic map	No. of BACs
Peach AFLP	AT/CAC		1
Peach AFLP	AT/CTA		4
Peach AFLP	EAA/MCAT10	Rootstock	1
Peach AFLP	TT/CCA		3
Peach cDNA	pP3Mi22		4
Peach RFLP	B2D4	WVa	1
Peach RFLP	B4A9	WVa	9
Peach RFLP	B4F12	WVa	4
Peach RFLP	B4G10	WVa	1
Peach RFLP	B6H11	WVa	3
Peach RFLP	B7D1	WVa	3
Peach RFLP	B7H6	WVa	1
Peach RFLP	B8A3	WVa	10
Prunus RFLP	AC10	Prunus	4
Prunus RFLP	AC13	Prunus	3
Prunus RFLP	AC31	Prunus	4
Prunus RFLP	AC33	Prunus	4
Prunus RFLP	AC37	Prunus	14
Prunus RFLP	AC43	Prunus	7
Prunus RFLP	AC55	Prunus	2
Prunus RFLP	AG2	Prunus	5
Prunus RFLP	AG12	Prunus	11
Prunus RFLP	AG33	Prunus	13
Prunus RFLP	AG35	Prunus	6
Prunus RFLP	AG37	Prunus	5
Prunus RFLP	AG43	Prunus	5
Prunus RFLP	AG44	Prunus	1
Prunus RFLP	AG56	Prunus	1
Prunus RFLP	AG105	Prunus	7
Prunus RFLP	OLE	Prunus	5
Prunus RFLP	Pij	Prunus	4
Prunus RFLP	Pru1	Prunus	2
Tomato cDNA	pTC11		4

Distribution of SSRs within BAC clones

SSR distribution was examined in sequences of three BAC clones. In each case, no effort was made to identify SSR-containing regions prior to sequencing. First, peach BAC clone PpN31C7 was sequenced in its entirety. This BAC was one of four overlapping BAC clones that hybridized with tomato clone pTC11, a cDNA from a gene-rich region of tomato (Mao et al. 2001). The 48,844-bp genomic sequence of PpN31C7 contained 22 SSR loci: 12 (T_n/A_n), nine $[(CT)_n/(GA)_n]$ and one $[(TA)_n/(AT)_n]$, where n is greater than or equal to ten. The sequence also contained nine regions with highly significant (E-values less than 1e-18) BLASTX similarity (i.e., similarity found by searching protein databases with conceptual translations of the query sequence in all possible reading frames) to sequences in *Arabidopsis thaliana* and other species. As shown in Fig. 4, likely gene sequences are concentrated in one half (0–25 kb) of the sequence and SSRs are concentrated in the other half (25–48.4 kb). The two putative gene sequences in the latter half were similar to reverse transcriptase (2e-15) and pol polyprotein (1e-18), respectively. To determine if these results were unique to this one region, two additional BAC clones were subjected to partial sequencing. One was selected from a contig that hybridized with peach cDNA probe pP3Mi2, thus we knew that this region contained genes. In an approximately 8,600-bp combined (non-contiguous) sequence, one compound repeat $[(CT)_{21}(CA)_{18}]$ and one interrupted repeat $[(TTA)_{12}N_{26}(TTA)_{13}]$ were identified. The final BAC

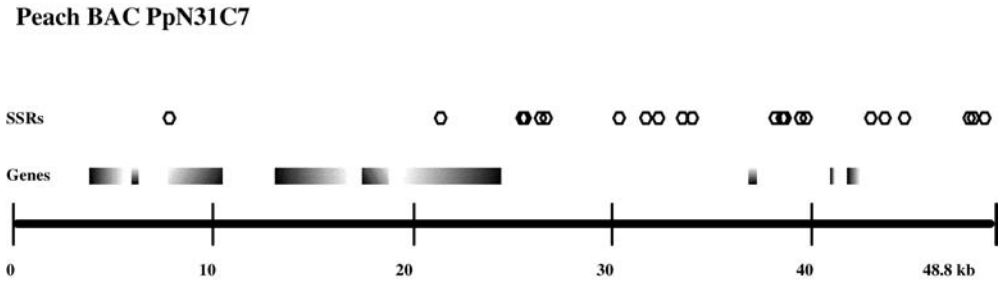

Fig. 4 Map of peach BAC clone PpN31C7. Putative gene regions are represented by *shaded boxes*; *open circles* denote SSRs

Table 3 SSRs derived from sequencing of Peach BAC clones

Name	Repeat type	Primers	Expected Nemared product (bp)	Polymorphism detected[a]
pchgms26	$(TTA)_{15}$	5′-TTTGATAGGATCCCAAGGGTA-3′ 5′-TTGGCTGGCAGTTATCATCA-3′	341	Yes
pchgms30	$(TA)_{14}$	5′-CATAAGGCTATTTGAGTGGCGCA-3′ 5′-CTTAGTTCCCCTATATTCCGTTTTC-3′	213	No[b]
pchgms31	$(CT)_{12}$	5′-TATCAGGTAAGGACCACTG-3′ 5′-GCTGCCGACGCTGTCAATTTC-3′	146	Yes
pchgms32	$(CT)_{21}$ $(CA)_{18}$	5′-TTGCTCTGTTGATTGGTGCT-3′ 5′-GATGAACTGCATGAGCGAAA-3′	195	No[b]

[a] In Peach rootstocks Lovell and Nemared

[b] Polymorphisms were detected in other peach rootstocks or varieties (A. Blenda, personal communication) or in Rose (L. Zhang, personal communication)

clone chosen hybridized with a peach AFLP probe EAA/MCAT10 and a priori we had no reason to suspect that this was a gene-containing region. In approximately 8 kb of sequence, one $(TTA)_{15}$ SSR was identified. Primers have been designed to amplify four of these SSR regions, and two detected differences between peach rootstocks Lovell and Nemared (Table 3).

Discussion

At the heart of any genomic resource is the existence of large insert clone libraries. However, published procedures for large insert library construction are technically challenging and expensive. To initiate the development of genomic resources in peach, we have constructed a useful, general purpose BAC library from peach DNA extracted and handled in solution. This method produced very clean, reasonably large DNA that, unlike DNA in nuclei embedded in agarose beads or plugs, was readily digested with very small amounts of restriction endonuclease. Size selection on conventional agarose gels allowed the use of inexpensive, universally available apparatus, and greatly reduced the electrophoresis run time, at the cost of reduced ability to resolve very large fragments of DNA. This approach significantly streamlines BAC library construction and yields libraries having insert sizes adequate for most applications. In our case, the relatively small genome of peach enabled good sequence representation (8-fold coverage) with a reasonable number of clones (44,000) of modest length (60 kb). For comparison, the peach scion BAC library of Wang et al. (2001) contains 21,000 clones with an average insert size of 95 kb for an expected 7-fold coverage of the genome.

In light of the utility and need of SSR markers, one of the first steps taken to characterize the new BAC library involved SSRs. Our previous estimates of the frequency of SSRs in peach involved screening non-enriched cDNA and small-insert genomic libraries (Sosinski et al. 2000). These estimates reflect the overall abundance of SSRs in peach, but necessarily provide no information on their genomic distribution. Overall, CT/GA repeats were estimated to occur every 100 kb, CA/GT repeats every 420 kb, and AGG/TCC repeats every 700 kb in the peach genome. Cipriani et al. (1999) identified SSRs in the peach genome utilizing enriched genomic libraries, so no estimate of SSR frequency was possible from their results; however, they also found that CT/GA repeats were more common than CA/GT repeats.

Having a large insert BAC library enabled us to evaluate SSR abundance and distribution in localized genomic regions of peach. Colony hybridizations of randomly chosen BACs suggested that SSRs occur with a frequency of at least one CT/GA locus every 78 kb, at least one CA/GT locus every 156 kb, and at least one AGG/TCC locus every 312 kb, assuming an average BAC insert size of 60 kb. This is significantly higher than initially estimated from plasmid library hybridizations (Sosinski et al. 2000). However, these averages can be misleading, as both Southern hybridization and sequence analysis provide evidence that individual BACs may contain different classes of repeat and more than one unit of a given repeat class, and thus that the repeat units may be clustered in the genome.

Further evidence for the density and clustering of SSR loci in peach was obtained by sequencing a gene-containing region (BAC clone PpN31C7). Mononucleotide T/A runs were found at a frequency of one per 4 kb of sequence; one CT/GA repeat locus was discovered per 5.4 kb of sequence, and one TA/AT locus per 49 kb of sequence. These values represent the overall density of SSRs in this clone; however, all but two of the SSRs were found in a region of the clone with a lower density of putative genes. This region also appeared to contain the remains of a retrotransposon (fragments of genes for reverse transcriptase and pol polyprotein). The abundance and distribution of SSRs in this peach genomic region resembles that in other plant species. In *A. thaliana*, almost one-third of all SSRs were of the mononucleotide T/A class, and almost two-thirds of all SSRs were found in intergenic regions (Cardle et al. 2000). A high percentage of barley SSRs were found to be associated with retrotransposons and other dispersed repetitive elements (Ramsay et al. 1999). Likewise, many spruce SSRs appear to be embedded in repetitive DNA (Morgante et al. 1996); however, these, like long SSRs associated with centromeric repeats in tomato (Areshchenkova and Ganal 1999), appear less likely to be suitable for genetic mapping.

Having a peach BAC library enabled us to begin developing a well-characterized genomic resource for the Rosaceae. Primers were designed to amplify four of the SSRs that were identified by sequencing BAC clones, and these loci are currently being utilized as markers in Rosaceae programs worldwide. When placed on genetic maps, these markers will facilitate alignment of these maps with physical map contigs from the BAC library described in this study. Likewise, a range of peach cDNA, RFLP and AFLP markers, many already placed on genetic maps, were readily assigned to physical map contigs (Table 2). These markers can serve as anchors for integrating genetic and physical genomic resources.

In summary, as we initiate development of integrated genomic resources for Rosaceae, we developed a simplified, cost-effective method for constructing BAC libraries, particularly appropriate for plant species of relatively minor economic importance. The results of initial studies with a peach BAC library provided information about the overall abundance and distribution of SSR markers; identification of SSR loci by hybridization or sequencing of randomly selected BAC clones; and the assignment of previously developed molecular markers (many already placed on genetic maps) to physical map contigs. This information is already being used to integrate rosaceous maps, identify genomic regions containing genetic loci for economically important traits, and develop polymorphic SSR markers for future mapping and genotyping projects. Thus, such peach BAC libraries can be used to further characterize the genomes of Rosaceae crop species, increasing our understanding of these important "minor" crops and our ability to wisely utilize their genomic resources for crop improvement.

Acknowledgements This work was supported by USDA CSREES Special Research Grant "Peach Tree Short Life in South Carolina", and by the South Carolina Agriculture and Forestry Research System under project SC-1700120. We thank Long Mao and Rod Wing for tomato probes pTC11 and pTC34; and Rebecca Ackerman for sequencing. We particularly thank Anne Westman for invaluable contributions to the preparation of this manuscript. The sequence of PpN31C7 has been deposited in GenBank under accession number AF467900. Technical Contribution No. 4724 of the SCAFRS.

References

Areshchenkova T, Ganal MW (1999) Long tomato microsatellites are predominantly associated with centromeric regions. Genome 42:536–544

Baird WV, Estager AS, Wells J (1994) Estimating nuclear DNA content in peach and related diploid species using laser flow cytometry and DNA hybridization. J Am Soc Hort Sci 119:1312–1316

Cardle L, Ramsay L, Milbourne D, Macaulay M, Marshall D, Waugh R (2000) Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics 156:847–854

Chaparro JX, Werner D, O'Malley D, Sederoff RR (1994) Targeted mapping and linkage analysis of morphological, isoenzyme, and RAPD markers in peach. Theor Appl Genet 87:805–815

Choi S, Creelman RA, Mullet JE, Wing RA (1995) Construction and characterization of a bacterial artificial chromosome library of *Arabidopsis thaliana*. Plant Mol Biol Rep 13:124–128

Cipriani G, Lot G, Huang W-G, Marrazzo MT, Peterlunger E, Testolin R (1999) AC/GT and AG/CT microsatellite repeats in peach [*Prunus persica* (L) Batsch]: isolation, characterisation and cross-species amplification in *Prunus*. Theor Appl Genet 99:65–72

Davis TM, Yu H (1997) A linkage map of the diploid strawberrry, *Fragaria vesca*. J Hered 88:215–221

Delseny M, Salses J, Cooke R, Sallaud C, Regad F, Lagoda P, Guideroni E, Ventelon M, Brugidou C, Ghesquière A (2001) Rice genomics: present and future. Plant Physiol Biochem 39:323–334

Dirlewanger E, Bodo C (1994) Molecular genetic mapping of peach. Euphytica 77:101–103

Dirlewanger E, Pronier V, Parvey C, Rothan C, Guye A, Monet R (1998) Genetic linkage map of peach [*Prunus persica* (L.) Batsch.] using morphological and molecular markers. Theor Appl Genet 97:888–895

Feinberg AP, Vogelstein B (1983) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132:6–13

Foolad MR, Arulsekar S, Becerra V, Bliss FA (1995) A genetic linkage map of *Prunus* based on an interspecific cross between peach and almond. Theor Appl Genet 91:262–269

Hemmat M, Weeden NF, Manganaris AG, Lawson DM (1994) Molecular-marker linkage map for apple. J Hered 85:4–11

Hurtado Ruiz M (1999) Obtención de marcadores moleculares para su aplicación a la mejora genética del albaricoquero. Tesis Doctoral, Instituto Valenciano de Investigaciones Agrarias

Joobeur T, Viruel MA, de Vicente MC, Jauregui B, Ballester J, Dettori MT, Verde I, Truco MJ, Messeguer R, Batlle I, Quarta R, Dirlwanger E, Arús P (1998) Construction of a saturated linkage map for *Prunus* using an almond × peach F2 progeny. Theor Appl Genet 97:1034–1041

Joobeur T, Periam N, de Vicente MC, King GJ, Arús P (2000) Development of a second generation linkage map for almond using RAPD and SSR markers. Genome 43:649–655

Lohuis MR ten, Meyer A, Meyer P (1993) An improved method for the isolation of high-molecular-weight >300 kb plant DNA in liquid phase. BioTechniques 14:890–892

Lu Z-X, Sosinski B, Reighard G, Baird WV, Abbott AG (1998) Construction of a genetic linkage map and identification of AFLP markers for resistance to root-knot nematodes in peach rootstocks. Genome 41:199–207

Lu Z-X, Sossey-Alaoui K, Reighard GL, Baird WV, Abbott AG (1999) Development and characterization of a codominant marker linked to root-knot nematode resistance and its application to peach rootstock breeding. Theor Appl Genet 99:115–122

Maliepaard C, Alston FH, van Arkel G, Brown LM, Chevreau E, Dunemann F, Evans KM, Gardiner S, Guilford P, van Huesden AW (1998) Aligning male and female linkage maps of apple (*Malus pumila* Mill.) using multi-allelic markers. Theor Appl Genet 97:60–73

Mao L, Begum D, Goff SA, Wing RA (2001) Sequence and analysis of the tomato *JOINTLESS* locus. Plant Physiol 126:1331–1340

Morgante M, Pfeiffer A, Costacurta A, Olivieri AM, Powell W, Vendramin GG, Rafalski JA (1996) Polymorphic simple sequence repeats in nuclear and chloroplast genomes: applications to the population genetics of trees. In: Ahuja MR, Boerjan W, Neale DB (eds) Somatic cell genetics and molecular genetics of trees. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp 233–238

Okie WR, Weinberger JH (1996) Plums. In: Janick J, Moore JN (eds) Fruit breeding vol 1. Tree and tropical fruits. Wiley, New York, pp 559–607

Rajapakse S, Belthoff LE, He G, Estager AE, Scorza R, Verde I, Ballard RE, Baird WV, Callahan A, Monet R, Abbott AG (1995) Genetic linkage mapping in peach using morphological, RFLP and RAPD markers. Theor Appl Genet 90:503–510

Rajapakse S, Byrne DH, Zhang L, Anderson N, Arumuganathan K, Ballard RE (2001) Two genetic linkage maps of tetraploid roses. Theor Appl Genet 103:575–583

Ramsay L, Macaulay M, Cardle L, Morgante M, degli Ivanissevich S, Maestri E, Powell W, Waugh R (1999) Initimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley. Plant J 17:415–425

Rosen S, Skaletsky HJ (1998) Primer3. Code available at http://www-genome.wi.mit.edu/genome_software/other/primer3.html

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Shimada T, Yamamoto T, Hayama H, Yamaguchi M, Hayashi T (2000) A genetic linkage map constructed by using an intraspecific cross between peach cultivars. J Japan Soc Hort Sci 69:536–542

Sosinski B, Gannavarapu M, Hager LD, Beck E, King GJ, Ryder CD, Rajapakse S, Baird WV, Ballard RE, Abbott AG (2000) Characterization of microsatellite markers in peach [*Prunus persica* (L.) Batsch]. Theor Appl Genet 101:421–428

Viruel MA, Messeguer R, de Vicente MC, Garcia-Mas J, Puidomènech P, Vargas F, Arús P (1995) A linkage map with RFLP and isozyme markers for almond. Theor Appl Genet 91:964–971

Wang D, Karle R, Brettin TS, Iezzoni AF (1998) Genetic linkage map in sour cherry using RFLP markers. Theor Appl Genet 97:1217–1224

Wang Q, Zhang K, Qu X, Jia J, Shi J, Jin D, Wang B (2001) Construction and characterization of a bacterial artificial chromosome library of peach. Theor Appl Genet 103:1174–1179